

P-ISSN: 2620 - 3383

Vol.3 No.2 Februari 2019

The logo consists of the letters 'TMD' in a bold, stylized, sans-serif font. The letters are green and have a thick, blocky appearance. The 'T' and 'M' are connected at the top, and the 'D' is also connected to the 'M'. The letters are set against a white background.

Technomedia Journal

iLearning Journal Center (iJC)

## **Evaluasi Performa Algoritma Decision Tree C4.5 Pada Dataset Bank Marketing Dengan Telepon Langsung**

Dikot Syarif Hidayatullah<sup>1</sup>  
Rahman Bayu Widhakdo<sup>2</sup>  
Ahmad Arivianto<sup>3</sup>  
Muhammad Lutfi Taqwim<sup>4</sup>  
Mochammad Yusa<sup>5</sup>

Akademisi Universitas Amikom Yogyakarta Fakultas Ilmu Komputer<sup>12345</sup>  
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta  
*E-mail: dikot.0400@students.amikom.ac.id*<sup>1</sup>, *rahman.w@students.amikom.ac.id*<sup>2</sup>,  
*ahmad.arivianto@students.amikom.ac.id*<sup>3</sup>, *muhammad.taqwim@students.amikom.ac.id*<sup>4</sup>,  
*mochammad.yusa@amikom.ac.id*<sup>5</sup>

### **ABSTRAK**

Marketing merupakan kegiatan untuk mempromosikan, menjual dan sebagai ujung tombak suatu perusahaan dalam menjalankan bisnis untuk memperoleh kesuksesan. Dalam hal ini marketing mencakup kegiatan sosialisasi dan juga pengiklanan dari suatu produk ke konsumen, terutama pada kasus bank yang memasarkan produk deposito berjangka ke nasabah dengan cara menelpon langsung nasabah. Begitu banyak cara bisa dikembangkan untuk menghasilkan sebuah hasil marketing yang efektif. Karena itu, diperlukannya evaluasi dari kegiatan marketing melalui telepon langsung untuk menentukan performa, knowledge dan rule berupa pohon keputusan dari proses marketing tersebut. Untuk memperoleh hasil evaluasi, maka diperlukan pengujian data mining klasifikasi dengan menggunakan Algoritma Decision Tree C4.5. Algoritma Decision Tree C4.5 dapat mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan dan memprediksi kelas target secara akurat dengan menggunakan variabel-variabel yang saling terkait [7]. Sehingga hasil evaluasi dapat digunakan sebagai bahan pertimbangan pihak bank dalam mengevaluasi tingkat keberhasilan teknik marketing dengan menelpon langsung pelanggan.

Kata kunci: Marketing, Klasifikasi, Algoritma Decision Tree C4.5, Bank

### **ABSTRACT**

*Marketing is an activity to promote, sell and spearhead a company in running a business to get success. In this case marketing includes socialization activities and also advertising from a product to consumers, especially in the case of the Bank that markets time deposit products to customers by calling customers directly. So many ways can be developed to produce an*

*effective marketing outcome. Therefore, the need for an evaluation of direct telephone marketing activities to determine the performance, knowledge and rule in the form of a decision tree from the marketing process. To obtain evaluation results, data mining testing classification is needed by using the Decision Tree C4.5 Algorithm. The Decision Tree C4.5 algorithm can change very large facts into decision trees that represent rules and accurately predict target classes using interrelated variables [7]. So the evaluation results can be used as material for consideration by the bank in evaluating the success rate of marketing techniques by calling customers directly.*

*Keywords: Marketing, Classification, Algorithm Decision Tree C4.5, Bank*

## **PENDAHULUAN**

*Marketing* atau pemasaran merupakan kegiatan untuk mempromosikan, menjual dan sebagai ujung tombak suatu perusahaan dalam menjalankan bisnis nya untuk memperoleh keuntungan dan kesuksesan. Ketika kegiatan marketing terlaksana dengan baik, maka peluang bisnis/usaha tersebut mempunyai peluang untuk sukses akan terbuka dengan lebar. Dalam hal ini *marketing* mencakup kegiatan pengenalan dan juga pengiklanan dari suatu produk ke konsumen, terutama pada kasus Bank memasarkan produk deposito berjangka ke nasabah dengan cara menelpon langsung pelanggan. Sedangkan sekarang model pemasaran semakin banyak variasinya. Begitu banyak cara bisa dikembangkan untuk menghasilkan sebuah hasil maksimal pemasaran yang efektif. Salah satu metode yang dipergunakan saat ini adalah menggunakan telepon. Terlebih lagi, ketika bisnis kita adalah sebuah bisnis yang lebih terfokus kepada pendekatan yang personal kepada calon konsumen.

Karena itu, diperlukannya evaluasi dari kegiatan marketing melalui telepon langsung untuk menentukan performa, *knowledge* dan *rule* atau aturan berupa pohon keputusan dari proses marketing tersebut. Untuk memperoleh evaluasi tadi maka diperlukan pengujian data mining klasifikasi dengan menggunakan algoritma *Decision Tree C4.5* pada Dataset terkait dengan kampanye marketing/pemasaran langsung dari sebuah lembaga perbankan Portugis. Pemasaran didasarkan pada panggilan telepon, dengan target dataset jika produk (deposito berjangka bank) akan berlangganan maka ('yes') atau jika tidak maka ('no') tidak berlangganan.

Teknik Klasifikasi merupakan salah satu teknik pada data mining. Tujuan atau objektif dari teknik klasifikasi data mining dengan Algoritma *Decision Tree C4.5* adalah untuk mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan dan memprediksi kelas target secara akurat dengan menggunakan variabel-variabel terkait [7]. Terdapat banyak model algoritma dalam Teknik klasifikasi data mining, Tetapi model algoritma klasifikasi dengan menggunakan Algoritma *Decision Tree C4.5* memiliki tingkat akurasi yang tinggi, hasil yang mudah dipahami dengan bahasa alami dan terkenal diantara algoritma lainnya [3]. Sehingga dengan melakukan penelitian ini dapat menghasilkan *output* berupa performa, *knowledge* dan *rule* berupa pohon keputusan dari proses marketing, sebagai pertimbangan pihak bank dalam mengevaluasi tingkat keberhasilan teknik marketing dengan menelpon langsung pelanggan.

## PERMASALAHAN

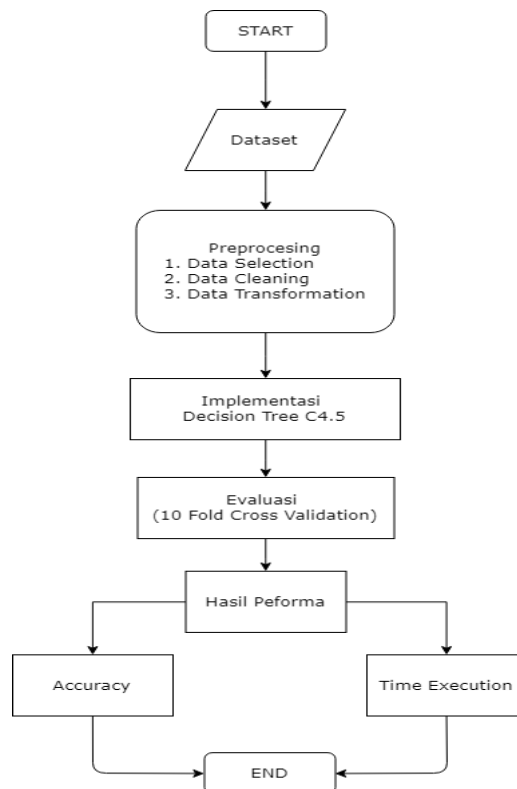
Aktifitas marketing atau pemasaran dalam sebuah bisnis merupakan ujung tombak yang menentukan kelangsungan jalannya bisnis. Dengan pendekatan marketing yang tepat, akan membuat produk/jasa tersosialisasi ke calon konsumen, meningkatkan volume penjualan, dan ujung-ujungnya menghasilkan profit. Tidak heran jika para pelaku usaha sangat concern dengan masalah marketing ini. Berbagai upaya dilakukan untuk meng-upgrade skill dan knowledge marketingnya, mengikuti berbagai pelatihan, memperluas networking, sampai ikut terjun langsung dalam kegiatan promosinya.

Namun, kesulitan dalam pemasaran juga sering kali terjadi pada saat menentukan jenis marketing yang bisa dikembangkan untuk menghasilkan sebuah hasil maksimal pemasaran yang efektif. Untuk studi kasus pada Instansi Bank, salah satu metode yang dipergunakan saat ini adalah menggunakan jasa telepon langsung ke pelanggan untuk melakukan pemasaran. Terlebih lagi, ketika bisnis kita adalah sebuah bisnis yang lebih terfokus kepada pendekatan yang personal kepada calon konsumen.

Penelitian ini menggunakan data yang berasal dari *UCI Machine Learning* yang dapat diunduh untuk penelitian-penelitian tentang data mining. Data penelitian ini berasal dari Lembaga Perbankan Portugis. Data ini bersifat real yang memang berasal dari gabungan-gabungan bank yang telah melakukan kampanye atau promosi pemasaran yang didasarkan pada panggilan telepon. Seringkali, lebih dari satu kontak ke klien yang sama diperlukan, untuk mengakses jika produk (deposito berjangka bank) akan ('ya') atau tidak ('tidak') berlangganan. Dataset yang digunakan untuk evaluasi pada penelitian ini terdiri dari 17 atribut dengan 4521 records. Atribut-atribut tersebut adalah *age, job, marital, education, default, balance, housing, loan, contact, day, month, duration, campaign, pdays, previous, poutcome, y*. Dataset penelitian (Bank Marketing) tersedia dan dapat diakses di website <https://archive.ics.uci.edu/ml/machine-learning-databases/00222/>.

## METODOLOGI PENELITIAN

Dalam penelitian ini terdapat dua tahap. Tahap yang pertama adalah preparasi data dan yang kedua adalah tahap evaluasi.



**Gambar 1.** Flowchart tahap penelitian

Pada Gambar 1 merupakan *Flowchart* tahapan penelitian dan proses data mining dengan menggunakan software Rapidminer versi 9.1 dan menggunakan komputer yang memiliki spesifikasi Processor AMD Ryzen 5-2500U, RAM 8GB, memori penyimpanan SSD 256GB. Berikut adalah tahapan-tahapan dan penjelasannya :

1. Tahap Preparasi Data

Data Preparation adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (input yang baik untuk data mining tools). Dikarenakan dataset yang tersedia berisikan informasi yang belum sesuai dengan proses data mining yang akan dijalankan, maka hal ini perlu dilakukan karena data yang biasanya berasal dari dunia nyata memiliki nilai atau data yang bersifat redundant, noisy, multiple (ganda), dan tidak komplit. Tahap-tahap yang akan dilakukan dalam preparasi data pada penelitian ini adalah sebagai berikut :

a. Data Selection

Tahap ini adalah langkah dimana akan dilakukannya tahap pemilihan atribut atau feature yang akan digunakan. Dalam tahap ini, penghapusan atribut yang tidak berpengaruh terhadap proses klasifikasi keberhasilan bank marketing dengan telepon langsung dihapuskan.

b. Data Cleaning

Proses analisa kualitas dari suatu data dengan cara mengubah, mengoreksi, atau menghapus data-data yang salah, tidak lengkap, tidak akurat, atau memiliki format yang salah dalam basis data agar dapat menghasilkan data berkualitas tinggi.

Dengan Menganalisis jumlah record yang tidak ada pada setiap atribut atau langkah identifikasi missing values dan menghindari terjadi data yang sama untuk setiap atribut.

c. Data Transformation

Langkah ini adalah mengkategorisasikan value-value yang bertipe numeric kemudian dirubah menjadi polinomial. Dan merubah ke interval value.

2. Tahap Evaluasi

Selanjutnya di jalankan proses evaluasi pada penelitian seperti pada Gambar 2. Pada skenario pengujian ini terdapat beberapa teknik pengujian sesuai dengan kapabilitas yang dimiliki oleh model algoritma tersebut. Setelah tahap *preprocessing* dilakukan, dataset di-import ke dalam *machine learning* Rapidminer 9.0. Skenario ujicoba pertama adalah dengan mengimplementasi model algoritma algoritma C4.5. Setelah dilakukan proses data mining selanjutnya akan dievaluasi menggunakan teknik *10-fold crossvalidation*. Metode Cross-validation (CV) ini adalah metode statistik yang dapat digunakan untuk mengevaluasi kinerja model atau algoritma dimana data dipisahkan menjadi dua subset yaitu data proses pembelajaran dan data validasi / evaluasi. Model atau algoritma dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi. Selanjutnya pemilihan jenis CV dapat didasarkan pada ukuran dataset. Biasanya CV K-fold digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi. Lalu hasilnya digenerating dengan *confusion matriks*. Setelah itu akan dihitung performanya berdasarkan performa akurasi dan *execution time*.

Berikut ini adalah beberapa literatur review terkait dengan penelitian yang akan dilakukan :

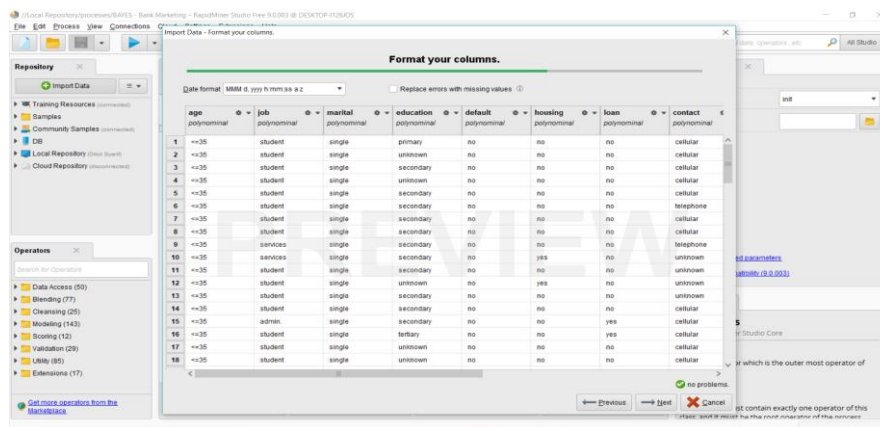
- a. Dari penelitian Oranuch Pantho dan Monchai Tiantong [1], diperoleh kesimpulan dengan menggunakan Algoritma *Decision tree C4.5* Dari banyak nya data di dapatkan *rule* yang simple dan spesifik.
- b. Dari penelitian Sri Wahyuni [2], Algoritma *Decision Tree C4.5* digunakan untuk klasifikasi, sehingga menghasilkan *knowledge* atau pengetahuan untuk mengetahui faktor-faktor sehingga terjadinya suatu kejadian tertentu. Dan Dalam pemilihan variabel data, yang merupakan atribut kondisi dan keputusan atribut, akan sangat mempengaruhi aturan atau pengetahuan yang dihasilkan. Semakin banyak data yang diproses, semakin banyak aturan atau pengetahuan yang didapat.
- c. Dari penelitian Khairul Sani [3], Pada analisis perbandingan Algoritma Klasifikasi untuk *Authentication* Uang Kertas, Setelah dilakukan uji coba menggunakan metode klasifikasi, maka persentase akurasi yang sangat tinggi yaitu menggunakan metode tree C4.5 dengan nilai akurasi sebesar 98,5 %, sedangkan neural network sebesar 95%, dan navies bayes sebesar 85%.



- d. Dari penelitian Teguh Budi Santoso [4], Algoritma *Decision Tree C4.5* digunakan untuk melakukan Prediksi Loyalitas Pelanggan dengan output hasil berupa pohon keputusan.
- e. Dari penelitian Rismayanti [5], Dalam memulai penelitian data mining menggunakan Algoritma *Decision Tree C4.5* perlu menjalankan beberapa tahap yakni : identifikasi masalah, menganalisa masalah, menentukan tujuan, mempelajari literatur, mengumpulkan data, mempersiapkan data dan implementasi.

## HASIL DAN PEMBAHASAN

1. Hasil Tahap Preparasi Data, pada Gambar 2 dilakukan preparasi data dengan melakukan import dataset bank marketing ke aplikasi RapidMiner. Selanjutnya menjalankan tahap-tahap preparasi data berikut :



**Gambar 2.** Tampilan Tahap Preparasi Data

- a. Data selection  
 Atribut-atribut yang dianggap tidak berpengaruh adalah education, contact, day, campaign, balance, pdays dan previous. Kemudian atribut tersebut dihilangkan dari dataset pada saat import dataset ke software Rapidminer.
- b. Data Cleaning  
 Pada dataset yang tersedia, dari hasil Analisa dengan aplikasi RapidMiner tidak ada missing value pada dataset dan tidak terdapat data yang duplikat disetiap atributnya dilihat dari result dataset pada Rapidminer.
- c. Data Transformation  
 Pada tahap ini yang dirubah atribut nya menjadi polynomial yaitu atribut age, campaign dan atribut duration dengan menggunakan operator discretize by entropy pada Rapidminer.

Dari proses preparasi data, dihasilkan dataset final untuk siap dilanjutkan keproses evaluasi. Berikut Tabel 2 merupakan hasil atribut dataset Bank Marketing Final setelah melalui tahap preparasi data.

**Tabel 1.** Daftar lengkap dataset final

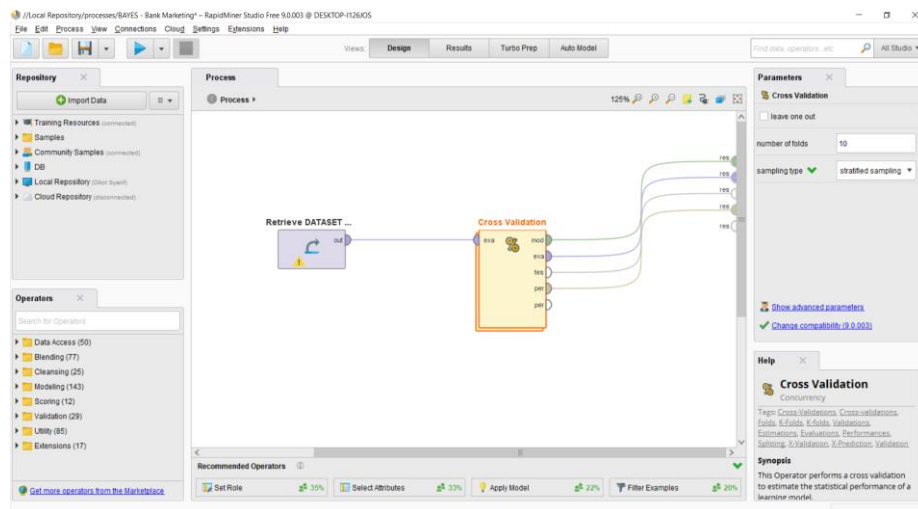
No	Nama Atribut	Tipe	Deskripsi
1	Age	Polynominal	Kategori age : $\leq 35$ , 46...65, $> 66$ Menunjukkan kategori usia pelanggan
2	Job	Polynominal	Kategori job : admin, blue collar, entrepreneur, housemaid, management, retired, self, employed, services, student, technician, unemployed, unknown. Merupakan Atirbut kategori pekerjaan pelanggan
3	Marital	Polynominal	Kategori Marital : divorced ,married, single, unknown. Merupakan kategori pernikahan pelanggan
4	Default	Polynominal	Kategori default : no ,yes ,unknown. Merupakan kategori apakah pelanggan memiliki kredit secara default
5	Housing	Polynominal	Kategori Housing : no, yes, unknown. Merupakan kategori kepemilikan rumah
6	Loan	Polynominal	Kategori Loan : : no, yes, unknown. Merupakan apakah pelanggan memiliki pinjaman pribadi
7	Month	Polynominal	Kategori Month : jan, feb, mar, ..., nov, dec. merupakan kategori bulan dimana dilakukan promosi ke pelanggan
8	Duration	Polynominal	Kategori Duration : range1 $[-\infty - 75]$ , range2 $[75 - 211]$ , range3 $[211 - 383]$ , range4 $[383 - 645]$ , range5 $[645 - \infty]$ . Merupakan kategori lama waktu atau durasi kontak terakhir
9	Poutcome	Polynominal	Kategori Poutcome : failure, nonexistent, success. Merupakan hasil dari kampanye



pemasaran sebelumnya

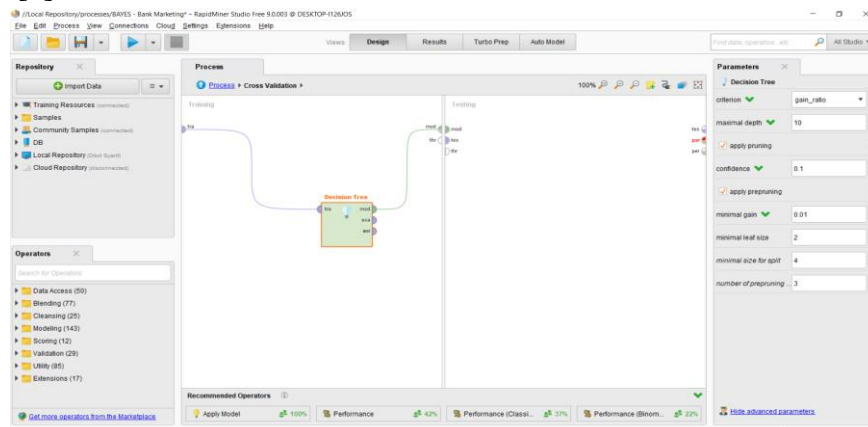
10 y      Polynominal      Kategori y : yes, no. Merupakan apakah klien  
 (Label      berlangganan deposito berjangka  
 target)

2. Tahap Evaluasi, pada tahapan evaluasi menggunakan operator *Cross Validation* pada proses evaluasi dengan menggunakan aplikasi RapidMiner.



**Gambar 3.** Tampilan Tahap Kedua Evaluasi

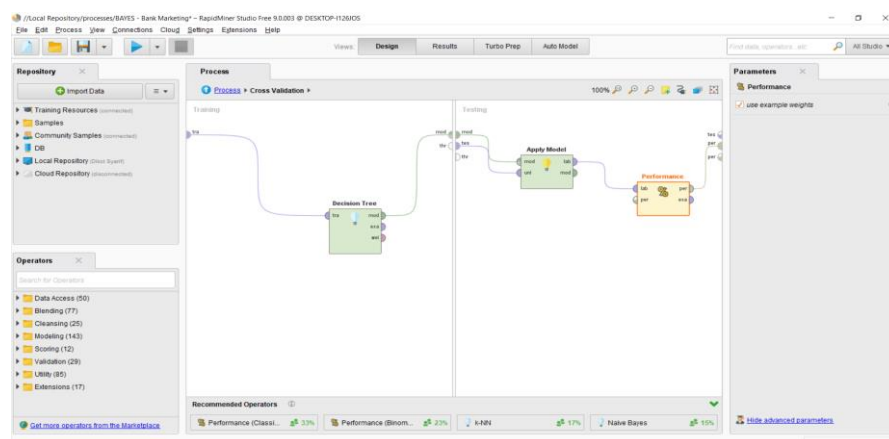
- a. Pada tahap evaluasi Gambar 3, peneliti menggunakan Cross Validation pada Rapidminer dengan parameter number of fold adalah 10 dan Sampling type adalah stratified\_sampling, Stratified Sampling membangun subset acak dan memastikan bahwa distribusi kelas dalam himpunan adalah sama seperti dalam ExampleSet seluruh [6].



**Gambar 4.** Tampilan Implementasi Algoritma Decision Tree

- b. Selanjutnya pada operator Cross-Validation memiliki port input yaitu, training example set (training) sebagai port input memperkirakan ExampleSet untuk melatih sebuah model (training data set). ExampleSet yang sama akan digunakan

- c. selama subproses pengujian untuk menguji model algoritma decision tree dengan menggunakan operator Decision Tree pada Gambar 4. Untuk membuat pohon keputusan, peneliti menggunakan operator Apply Model. Operator ini menerapkan suatu model terlatih pada sebuah ExampleSet. Sebuah model pertama kali dilatih di sebuah ExampleSet, informasi yang berkaitan dengan ExampleSet dipelajari oleh model. Maka model tersebut dapat diterapkan pada ExampleSet yang lain [6].



**Gambar 5.** Tampilan Menghitung Accuracy dengan Operator Performance

- d. Dalam membuat pohon keputusan untuk menentukan knowledge atau rule klasifikasi bank marketing dengan cara menelpon pelanggan secara langsung, kita juga menggunakan operator Performance [6]. Operator ini digunakan untuk evaluasi kinerja. Operator ini memberikan daftar nilai kriteria kinerja yaitu hasil performa berupa *accuracy*. Gambar 5 melakukan tahap untuk menghitung nilai *accuracy* performa dengan menggunakan operator *Performance*.

### 3. Hasil Performa

Di dalam penelitian, dua parameter pendekatan perhitungan performa algoritma klasifikasi diperhitungkan. Parameter-parameter tersebut adalah Akurasi (*accuracy*) dan *Execution time*. *Accuracy* dalam klasifikasi adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi. Rumus *Accuracy* dapat didefinisikan pada rumus (1) berikut ini:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

Dimana, TP merupakan jumlah *record/tuple* positif yang dilabelkan secara benar oleh model algoritma klasifikasi, TN merupakan jumlah *record/tuple* negatif yang dilabelkan secara salah oleh model algoritma klasifikasi, dan P merupakan total semua *record* yang dievaluasi. *Execution time* merupakan salah pengukuran algoritmik yang digunakan untuk mengukur efesiensi sebuah model algoritma

	true yes	true no	class precis
yes	10	1	0.91
no	1	10	0.91

Dari dari hasil pohon keputusan dan model aturan berbentuk teks diperoleh

knowledge atau rule klasifikasi bahwa, jika berdasarkan durasi telepon marketing bank lebih dari >645s (11 menit) antara pihak bank dan pelanggan dan pelanggan berstatus single, umur pelanggan <35, menunjukkan terdapat paling banyak pelanggan bank yang melakukan deposito berjangka. Sedangkan klasifikasi dengan durasi telepon <75s (75 detik) paling banyak menunjukkan pelanggan untuk tidak melakukan deposito ke bank. Jika berdasarkan waktu paling menurut klasifikasi bulan maret memiliki persentase tertinggi dibandingkan bulan lain nya. Jika berdasarkan kepemilikan rumah pelanggan yang memiliki rumah lebih banyak menerima tawaran bank, jika berdasarkan job pekerjaan admin punya potensi menerima tawaran bank lebih tinggi dibandingkan pekerjaan lain, dan berdasarkan dari upaya promosi bank sebelumnya yang memiliki kategori success punya potensi berlangganan lebih tinggi dibanding kategori lain nya.

## **KESIMPULAN**

Berdasarkan data hasil evaluasi kinerja algoritma klasifikasi dengan Decision Tree C4.5 untuk melakukan klasifikasi keberhasilan Bank Marketing dengan telepon langsung pelanggan dapat disimpulkan memiliki nilai performa 89,63% berdasarkan performa akurasi dan performa execution time nya 0s. Dari performa akurasi yang dihasilkan diperoleh knowledge bahwasannya tingkat keberhasilan bank marketing dimana pelanggan menerima tawaran dari bank di dasari dengan klasifikasi durasi telepon marketing bank lebih dari >645s (11 menit) antara pihak bank dan pelanggan dan pelanggan berstatus single, umur pelanggan <35, menunjukkan terdapat paling banyak pelanggan bank yang melakukan deposito berjangka. Sedangkan klasifikasi dengan durasi telepon <75s (75 detik) paling banyak menunjukkan pelanggan untuk tidak melakukan deposito ke bank. Jika berdasarkan waktu menurut klasifikasi, bulan maret memiliki tingkat keberhasilan marketing tertinggi dibandingkan bulan lain nya. Jika berdasarkan kepemilikan rumah pelanggan yang memiliki rumah lebih banyak menerima tawaran bank, jika berdasarkan job pekerjaan admin punya potensi menerima tawaran bank lebih tinggi dibandingkan pekerjaan lain, dan berdasarkan dari upaya promosi bank sebelumnya yang memiliki kategori success punya potensi berlangganan lebih tinggi dibanding kategori lain nya.

## **SARAN**

Berdasarkan dari pembahasan yang telah diuraikan sebelumnya, penulis dapat menyimpulkan saran sebagai berikut :

1. Pada penelitian ini, hasil evaluasi masih dapat dimaksimalkan dengan melakukan pengujian menggunakan algoritma klasifikasi lain selain Decision Tree C4.5.
2. Hasil data mining dengan menggunakan Decision Tree C4.5 dengan Jumlah data yang besar akan menghasilkan pohon keputusan dengan deskripsi yang panjang dimana masih terdapat kemungkinan penemuan aturan baru dari kasus diatas.

## DAFTAR PUSTAKA

- [1] Oranuch, P., & Monchai, T. (2016). *Using Decision Tree C4.5 Algorithm to Predict VARK Learning Styles*. International Journal of the Computer.
- [2] Wahyuni, S. (2018). *Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree*. Journal of Physics - IOP eBook Publishing.
- [3] Khairul, S., Winarno, W.W, & Fauziati, S. (2016). Analisis Perbandingan Algoritma *Classification* Untuk *Authentication* Uang Kertas. Jurnal Informatika.
- [4] Santoso, B.T. (2013). Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. Jurnal Ilmiah Fakultas Teknik LIMIT'S Vol. 10 No.1.
- [5] Rismayanti. (2016). Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan. Jurnal Media Infotama.
- [6] Aprilla, D. C., Baskoro, A.D., Lia, A., Wicaksana. I.W.S., Belajar Data Mining dengan RapidMiner. Academia.edu.
- [7] Kusrini & Lutfi, E.T. (2009). Algoritma Data Mining. Yogyakarta: Penerbit Andi.